DOCUMENT RESUME

ED 138 628 TM 006 251

AUTHOR Plake, Barbara S.; Hoover, H. D.
TITLE The Comparability of Equal Raw Scores for Children at Different Grade Levels: An Issue in "Out-of-Level" Testing.
PUB DATE [Apr 77]
NOTE 20p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New York, New York, April 5-7, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Achievement Tests; Analysis of Variance; Elementary Education; Grade 4; Grade 6; Grade 7; Grade 8; *Instructional Program Divisions; Item Analysis; *Raw Scores
IDENTIFIERS Iowa Tests of Basic Skills; *Out of Level Testing

ABSTRACT

To determine if equal raw scores have the same "meaning" for students tested "in" and "out-of-level," four "out-of-level" grade groups who took the Vocabulary, Reading, and Mathematics Concepts subtests of the Iowa Tests of Basic Skills typically administered to fifth graders were compared to "in level" control groups. The analyses revealed a few significant item-by-group interactions, but the magnitude of these components suggested their influence was minimal in overall score variability. It was concluded that equal raw scores have essentially the same "meaning" regardless of a child's actual grade level. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED138628

TM006 251

THE COMPARABILITY OF EQUAL RAW SCORES FOR CHILDREN AT DIFFERENT GRADE LEVELS: AN ISSUE IN "OUT-OF-LEVEL" TESTING

Barbara S. Plake
The American College Testing Program

H. D. Hoover
The University of Iowa

Achievement tests are frequently designed in levels with each level consisting of a subset of items designed for the typical child at a particular grade level. When a student's development has been at a rate not typical of her/his peers, it is often desirable to administer a level of the achievement battery other than the one designed for the grade level in which the child is a member. This is referred to as "out-of-level" testing.

When equal raw scores are obtained on the same level of a given subtest of an achievement exam, the procedure followed by most test authors for assigning derived scores of the developmental type is to assign the same derived score to the equal raw scores, regardless of the actual grade level of the pupil. In other words, all children with the same number of correct answers on a given level of the exam receive the same developmental score.

The purpose of this study was to examine the comparability of equal raw scores on a given level of an achievement test when these raw scores were obtained by different grade children. It was decided that raw scores would be considered "equal" if the items on the test failed to function differentially for groups of in- and out-of-level students, that is, retain the same relative difficulty for both groups. Those items demonstrating relatively different difficulty were considered potentially biased. In a

between group ANOVA the existence of an items by groups interaction was used as a signal of the presence of potentially biased items.

The Iowa Tests of Basic Skills was the achievement battery selected for study. Due to its overlapping format, the ITBS is particularly amenable to out-of-level testing. Three of the seven subtests were identified for consideration: Vocabulary, Reading, and Mathematics Concepts. These three were chosen because of their varying degree of curriculum dependency.

The data were from a mid-year testing date in Iowa in 1974 and 1975. All the students included in this study were assigned level 11 of the ITBS, the level usually administered to fifth graders. The students in fifth grade composed the in-level pool of students. Four out-of-level pools were formed by students in grades 4, 6, 7, and 8 that were assigned the level 11 subtests.

For each subtest chosen for study, four pairings of in- and out-of-level students were considered. For each pairing, two groups were selected from the in-level and out-of-level pools in such a manner that the two groups were equal in size and had nearly equal distributions of raw scores. Each of these pairings and matched-group selection was done within school system. For each school system, the item difficulties were calculated for each item in the subset for the in- and out-of-level groups. These item difficulties were then subjected to an arcsin transformation to stabilize the variance to $\frac{1}{N_j}$, where N_j is the number of students in either of the groups in school system j.

An Analysis of Variance procedure was used to analyze the data. When a significant items by groups interaction was obtained a follow-up procedure

which examined the interaction contrast for each individual item was used to identify specific items which were contributing to this interaction. The percent of total variance accounted for by the items by groups component was also determined.

The Vocabulary subtest revealed three significant ($\alpha = .05$) items by groups interactions: for out-of-level grades 6, 7, and 8. An analysis of the interaction contrasts identified three items that contributed to the items by groups interactions, two of which were easier for the out-of-level groups (see Figures 1-4). Further investigation revealed that these two items contained vocabulary words more often found in material of the higher grades. The percent of variance accounted for by each items by groups component was so small (less than one percent) that the influence of the items by groups component in relation to overall variability was considered minimal.

For the Reading subtest only one pairing produced an items by groups interaction that was significant ($\alpha = .05$). This was found for the grade 8 with grade 5 comparison. As was the case with the Vocabulary subtest the percent of variance accounted for by this component was minimal (less than one percent) in relation to the overall design. The follow-up analysis on the interaction contrasts from the grade 8 with grade 5 comparison identified one item as a possible contributor to the items by groups interaction and it was systematically easier for the grade 5 children (see Figures 5-8).

All four out-of-level comparisons for the Mathematics Concepts subtest provided significant items by groups interactions. Several items were identified as contributors to their respective items by groups interactions (see Figures 9-12). As was found in the other subtests the percents of variance

accounted for by the items by groups interactions were inconsequential (less than one percent) in all four grade level comparisons.

Two additional analyses were carried out on the Mathematics Concepts data. First, three elementary math curriculum experts were asked to independently select items that were considered potentially biased from the level 11 test for the four grade levels considered in this study. Of the 42 items, at least two of the experts agreed on eight items as being potentially biased toward one of the groups being compared. Of these eight items, in two cases, the data demonstrated results opposite to the predicted. Only one of these eight items was also selected by the ANOVA procedures as a contributor to the items by groups interaction.

The second additional procedure was to reanalyze the Mathematics Concepts data, breaking the raw score distribution into five score bands: very high, high, middle, low, and very low. For each of the four grade levels investigated all of the students in both the in- and out-of-level groups were combined, regardless of system membership, and then divided into these five score bands. This analysis was performed to determine if an interaction of items by groups by score bands was present. Although two of the interaction of items by groups by score bands were marginally significant, the percents of variance accounted for by these terms indicated the influence of this triple interaction was minimal for each of the grade level comparisons considered.

Table 1 contains a summary of the percents of total variance accounted for by the items by groups components in each of the four pairings for the three subtests chosen for study. It is seen in Table 1 that in every case, the percent of variance accounted for by the items by groups component is less than 1 percent of the total variance.

From the results reported previously, several conclusions are possible:

1. For the Vocabulary subtest, three items were identified by the ANOVA procedures as potentially biased. Two of these items were easier for the students in grades 6, 7, and 8 than for the in-level fifth grade students. An investigation of these items revealed that the vocabulary level of the words were typically from higher grades than fifth grade.

2. The analysis of the Reading subtest data revealed one item that could be considered possibly biased for the in-level and out-of-level grade 8 comparison. This item was easier for the fifth graders than for students in the eighth grade group.

3. The Mathematics Concepts results showed significant items by groups interactions at all four grade levels investigated. As was expected, the items identified by the ANOVA procedures were relevant to typical curriculum placement at grade levels other than grade 5.

4. The experts' predictions of biased items from the Mathematics Concepts subtest identified eight "consensus" items. Of these eight, only one was also identified by the ANOVA procedures. In two cases, the data suggested opposite directional differences from the predicted.

5. When the Mathematics Concepts interaction of items by groups was further investigated for each grade level by the five score bands, two of the items by groups by score levels interactions were significant. The percent of variance accounted for by this component was minimal at each grade level; this suggests that the items by groups interaction functions consistently for students across the five score bands.

From the above conclusions, it seems clear that these subtests, on the whole, do not function much differently for out-of-level students in the four grade levels investigated. Equal scores received on a subtest of a given level of the ITBS appear to have essentially the same

"meaning" regardless of the in- or out-of-level status of the children taking the test.

Table 1

Percent of Total Variance Accounted For by the Estimated Variance Components for Vocabulary, Reading, and Mathematics Subtests

	Grade Comparison	Components				
		Systems	Items	Items by Groups	Items by Systems	Items by Groups by Systems
Vocabulary	4	83.8	4.0	0.01	6.9	5.2
	6	86.3	2.2	0.04	4.1	3.3
	7	92.1	1.7	0.08	3.14	3.0
	8	91.2	1.9	0.02	3.6	3.0
Reading	4	87.4	3.6	0.00	5.2	3.8
	6	91.3	2.8	0.005	3.1	2.7
	7	88.6	3.1	0.00	4.6	3.8
	8	82.9	8.6	0.14	4.6	3.7
Mathematics Concepts	4	85.8	1.2	0.2	7.0	5.8
	6	86.8	3.8	0.1	5.2	4.0
	7	90.9	1.8	0.1	3.8	3.3
	8	83.5	3.5	0.2	6.3	6.5

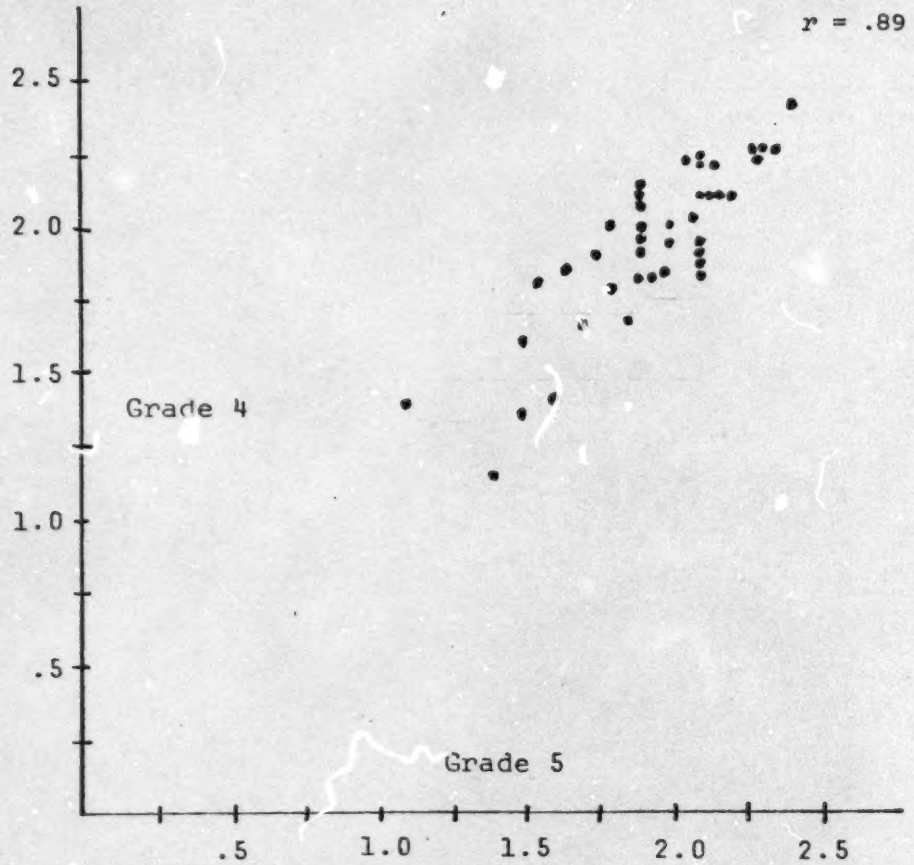

Figure 1. Vocabulary Subtest: Scatterplot of Arcsin Transformed Item Difficulties, Grade 4 vs. Grade 5

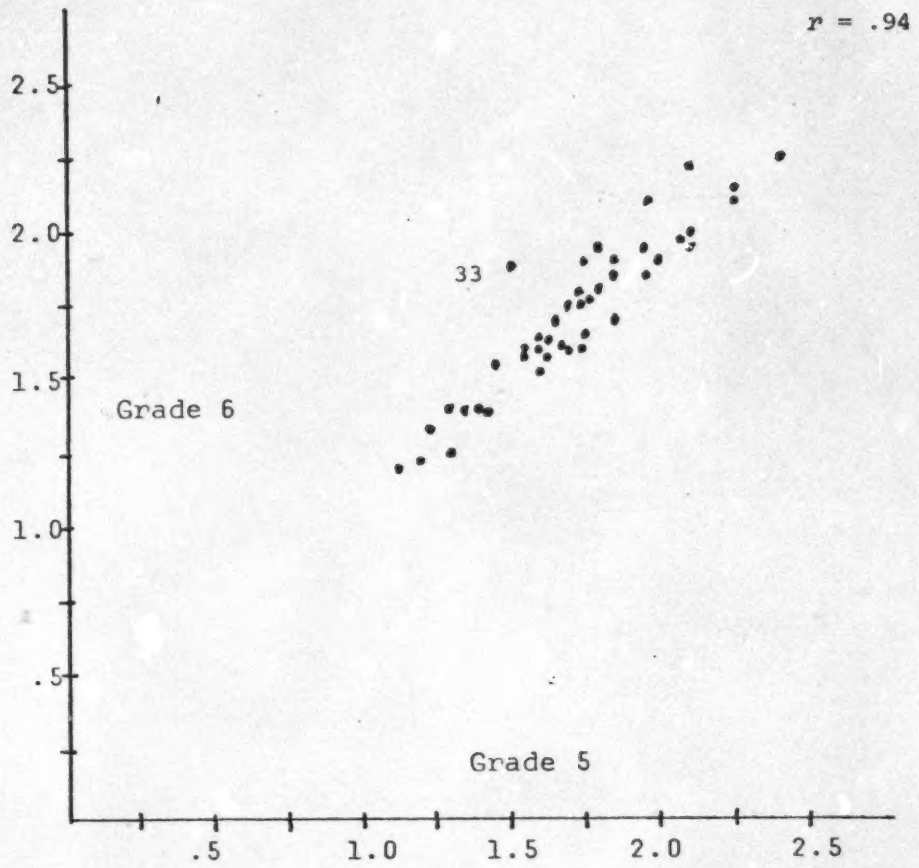

Figure 2. Vocabulary Subtest: Scatterplot of Arcsin Transformed Item Difficulties, Grade 6 vs. Grade 5

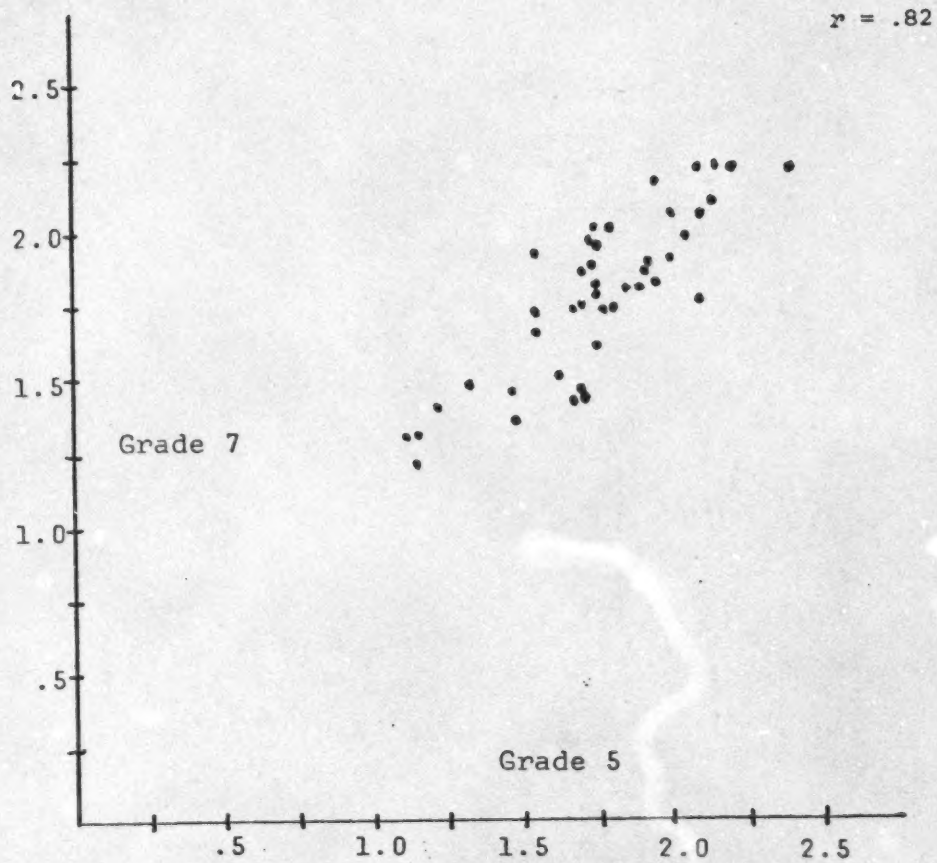

Figure 3. Vocabulary Subtest: Scatterplot of Arcsin Transformed Item Difficulties, Grade 7 vs. Grade 5

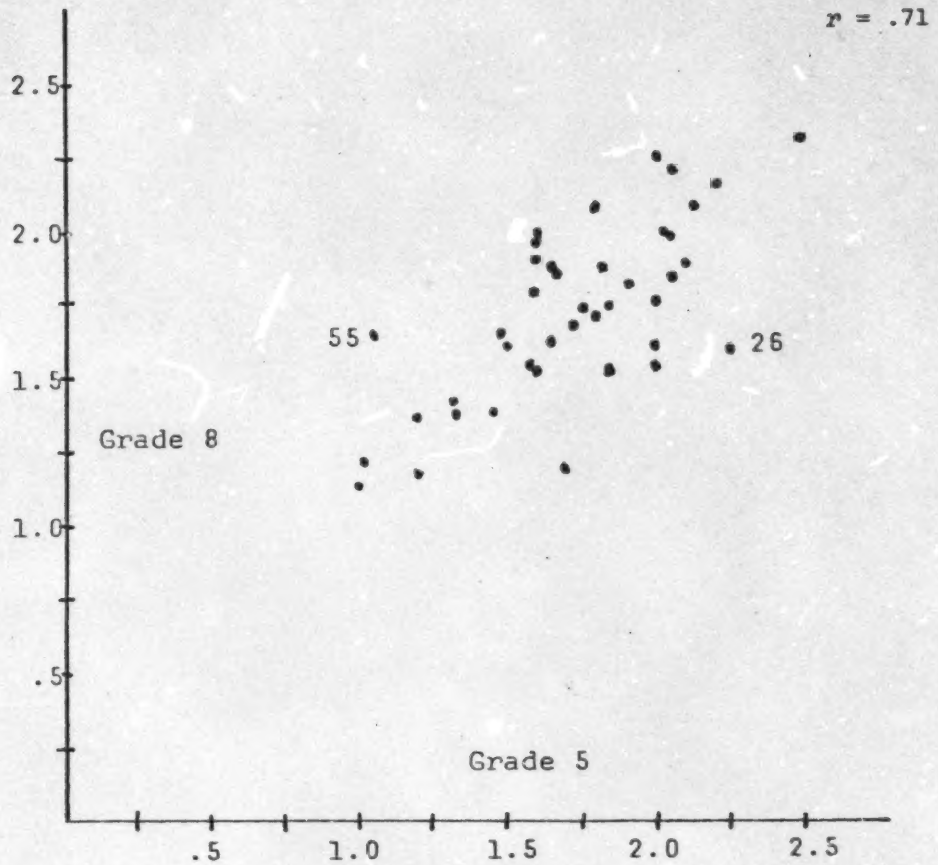

Figure 4. Vocabulary Subtest: Scatterplot of Arcsin Transformed Item Difficulties, Grade 8 vs. Grade 5

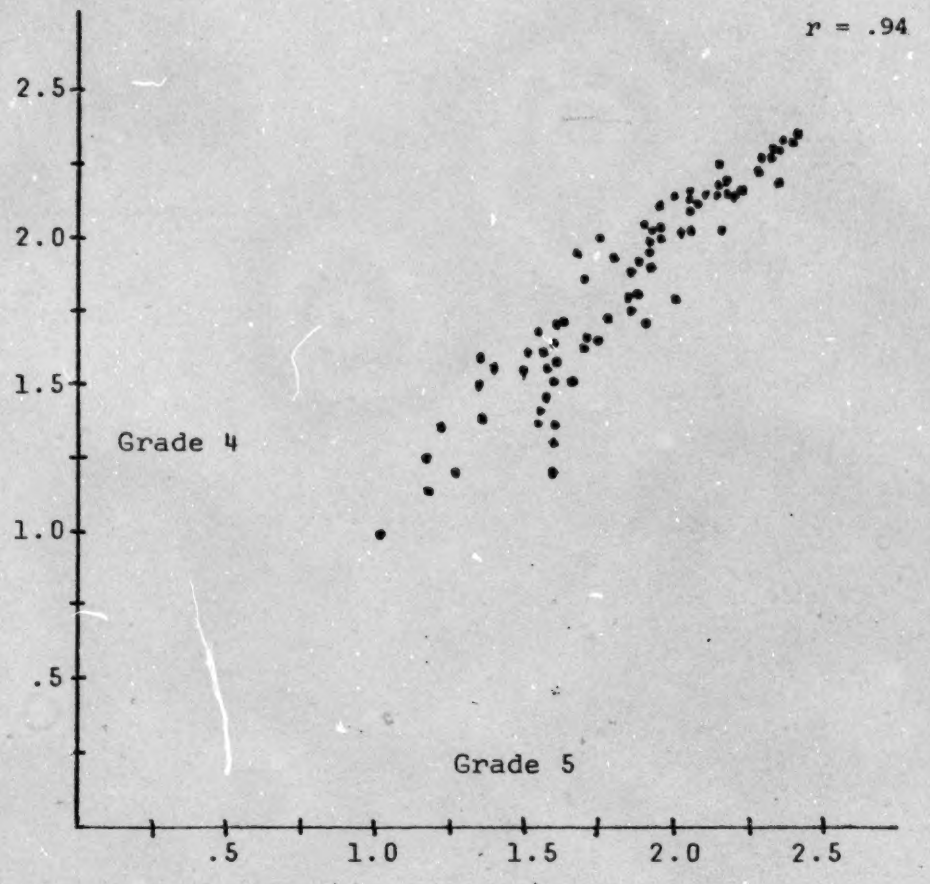

Figure 5. Reading Subtest: Scatterplot of Arcsin Transformed Item Difficulties, Grade 4 vs. Grade 5

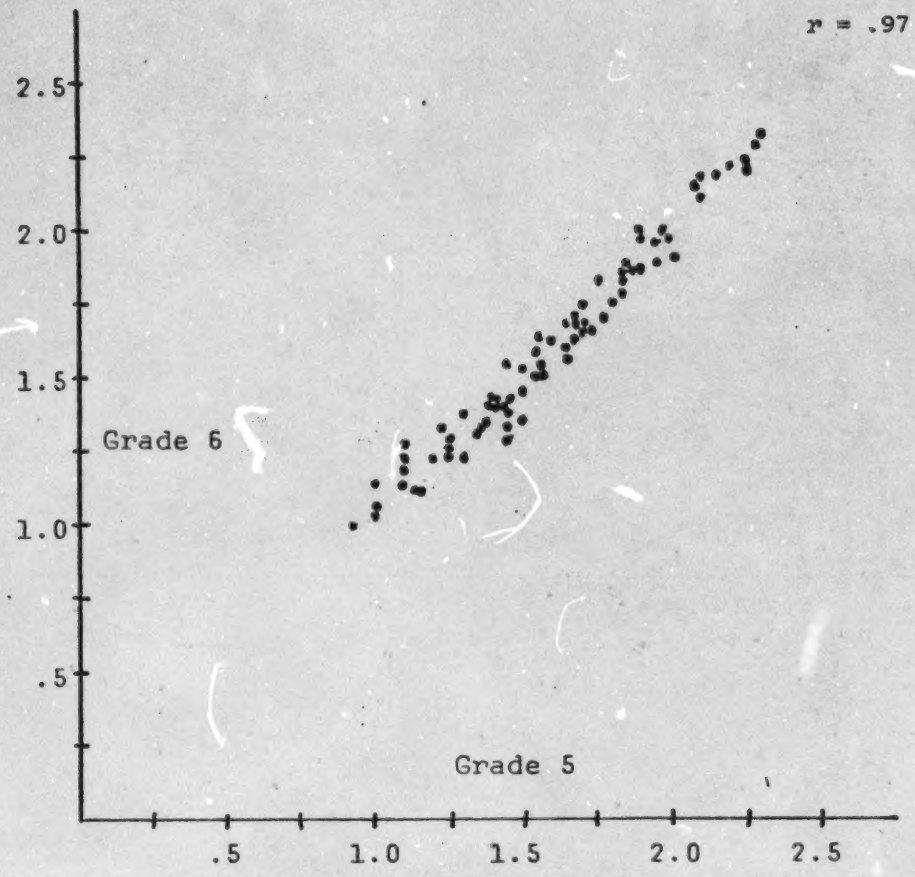

Figure 6. Reading Subtest: Scatterplot of Arcsin Transformed Item Difficulties, Grade 6 vs. Grade 5

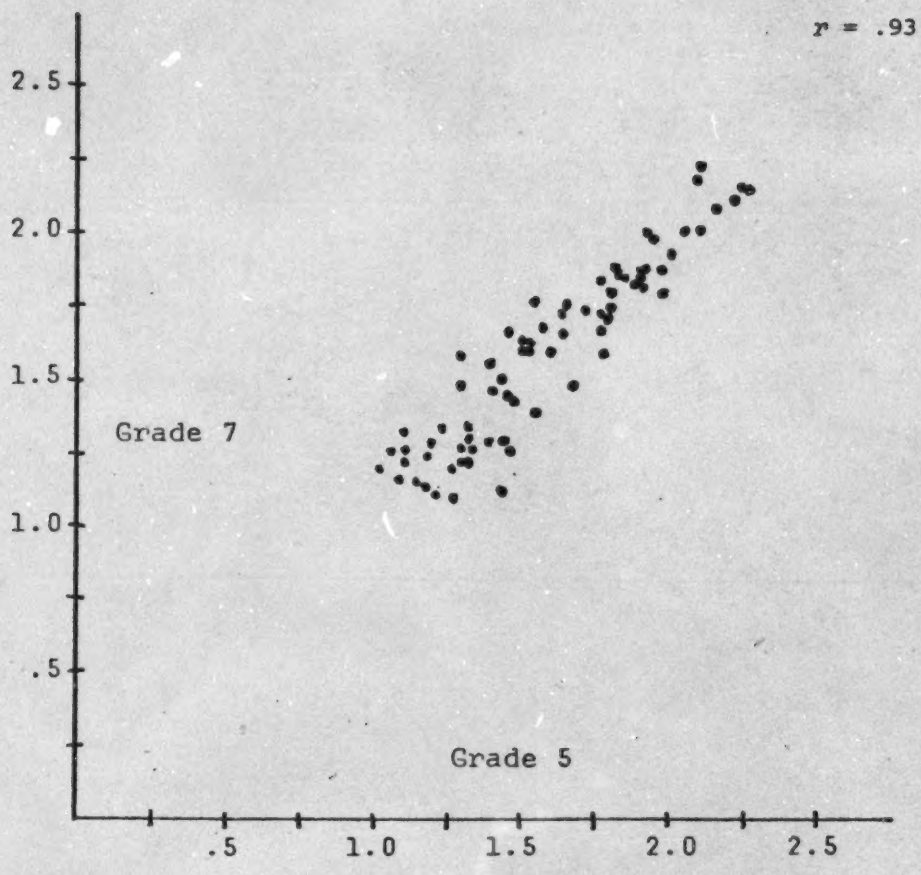

Figure 7. Reading Subtest: Scatterplot of Arcsin Transformed Item Difficulties, Grade 7 vs. Grade 5

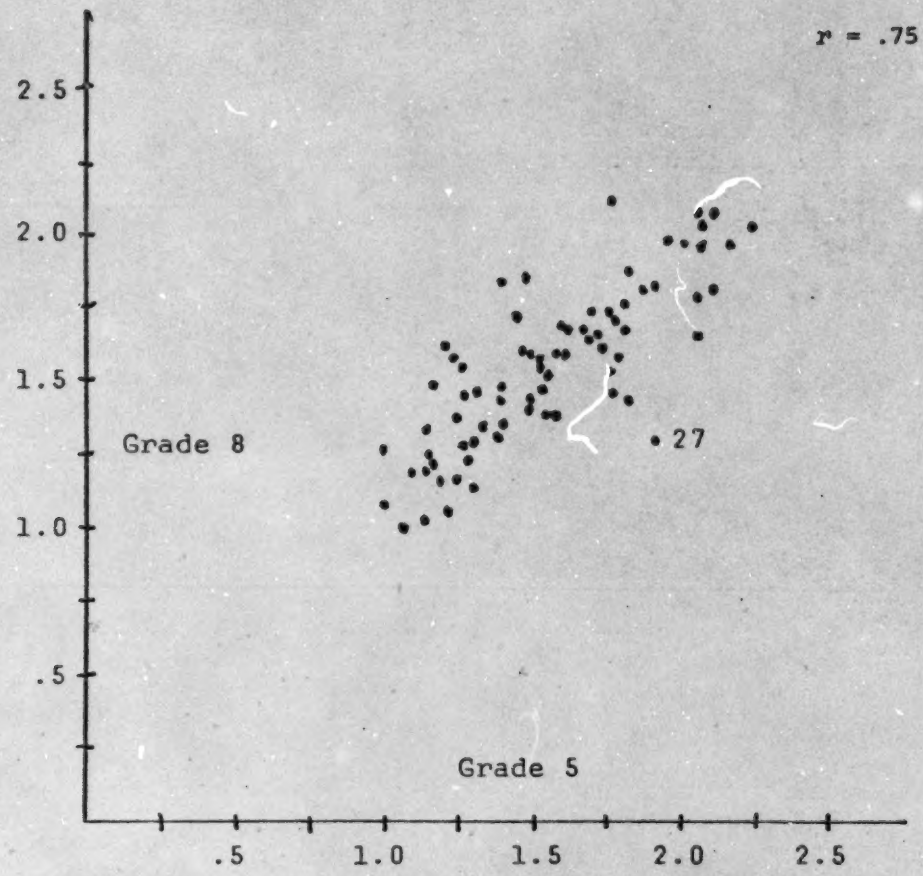

Figure 8. Reading Subtest: Scatterplot of Arcsin Transformed Item Difficulties, Grade 8 vs. Grade 5

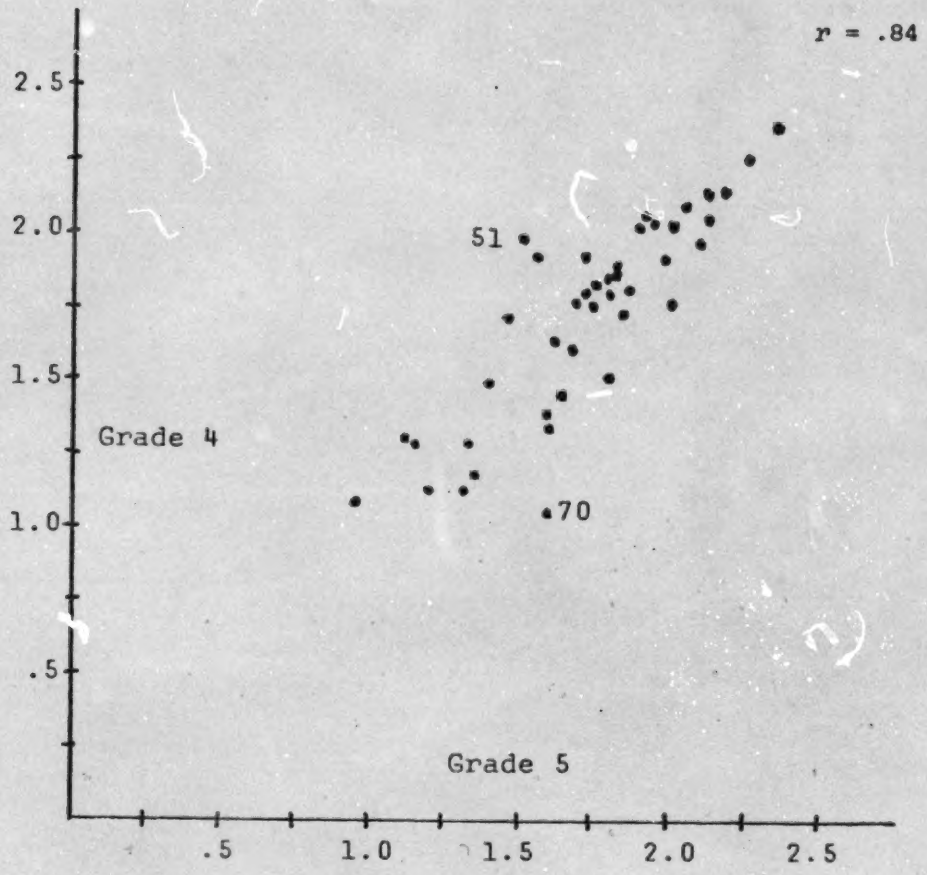

Figure 9. Mathematics Concepts Subtest: Scatterplot of Arcsin Transformed Item Difficulties, Grade 4 vs. Grade 5

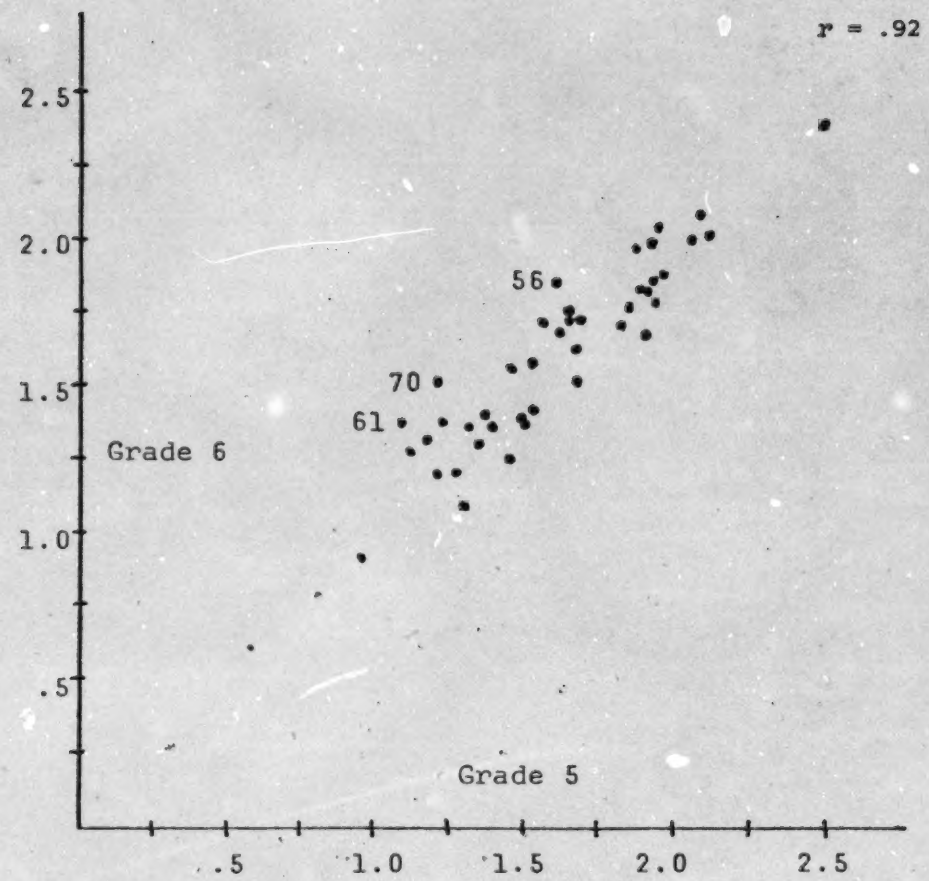

Figure 10. Mathematics Concepts Subtest: Scatterplot of Arcsin Transformed Item Difficulties, Grade 6 vs. Grade 5

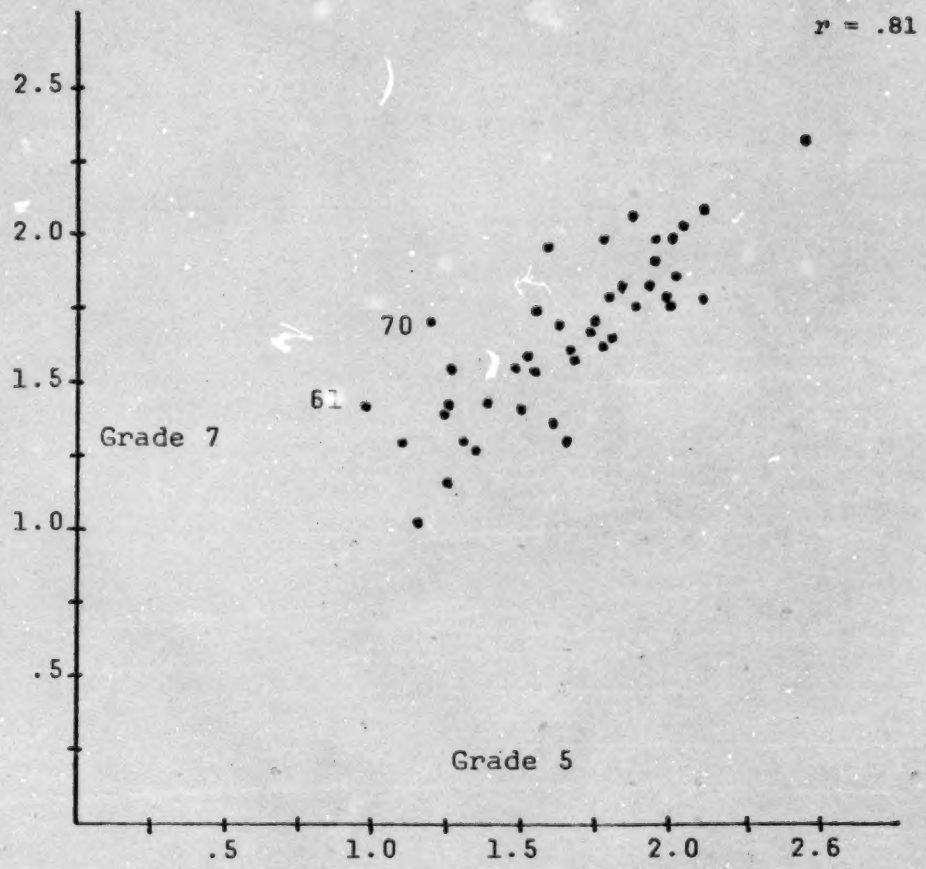

Figure 11. Mathematics Concepts Subtest: Scatterplot of Arcsin Transformed Item Difficulties, Grade 7 vs. Grade 5

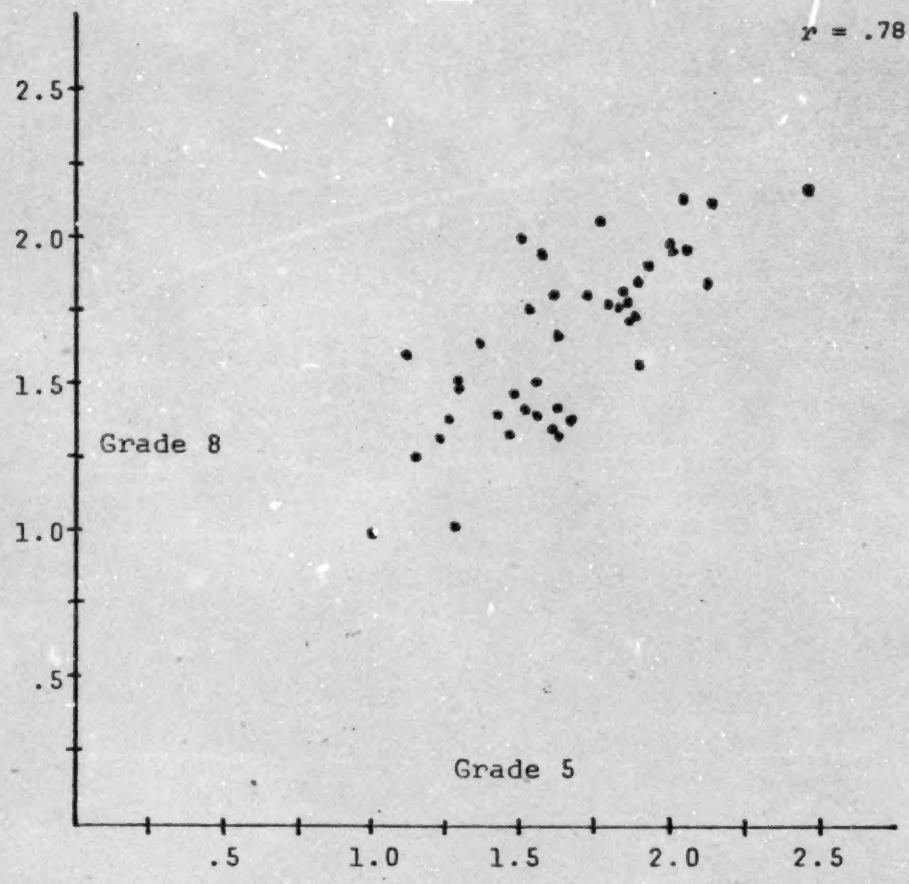

Figure 12. Mathematics Concepts Subtest: Scatterplot of Arcsin Transformed Item Difficulties, Grade 8 vs. Grade 5